

DIALIGN 2.1

Developed by Burkhard Morgenstern, Said Abdeddai
m, Kornelie Frech,
Klaus Hahn, Thomas Werner, Jens Stoye, Andreas D
ress

e-mail: burkhard.morgenstern@rp-rorer.co.uk

Published research assisted by DIALIGN 2 should
cite:

B. Morgenstern (1999),
"DIALIGN 2: improvement of the segment-to-segm
ent
approach to multiple sequence alignment."
Bioinformatics 15, 203 - 210.

Options:

=====

- 1) nucleic acid sequences aligned
- 2) no translation of of nucleotide diagonals into pepti
de diagonals
- 3) 5 "*" characters for regions of maximum similarity

Aligned sequences:

=====

length:

=====

- | | | |
|----|------------|------|
| 1) | EP | 2541 |
| 2) | 09/471,669 | 1503 |

Average sequence length: 2022.000

Please note that only upper-case letters are considered to
be aligned.

For more information, have a look at the user guide

http://bibiserv.techfak.uni-bielefeld.de/dialign/user_guide2.html

Alignment (DIALIGN format):

=====

```
EP      1  ATGGCCCAAG CCCTGCCCTG GCTCCTGCTG TGGATGGGC
G CGGGAGTGCT
09/471,669
G CGGGAGTGCT  1  ATGGCCCAAG CCCTGCCCTG GCTCCTGCTG TGGATGGGC
```

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

```
EP      51  GCCTGCCCCAC GGCACCCAGC ACGGCATCCG GCTGCCCCCT
G CGCAGCGGCC
09/471,669
G CGCAGCGGCC  51  GCCTGCCCCAC GGCACCCAGC ACGGCATCCG GCTGCCCCCT
```

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

EP 101 TGGGGGGCGC CCCCCTGGGG CTGCGGCTGC CCCGGGAGA
C CGACGAAGAG 09/471,669 101 TGGGGGGCGC CCCCCTGGGG CTGCGGCTGC CCCGGGAGA
C CGACGAAGAG

* *****
* *****
* *****
* *****
* *****

EP 151 CCCGAGGAGC CCGGCCGGAG GGGCAGCTTT GTGGAGATG
G TGGACAACCT 09/471,669 151 CCCGAGGAGC CCGGCCGGAG GGGCAGCTTT GTGGAGATG
G TGGACAACCT

* *****
* *****
* *****
* *****
* *****

EP 201 GAGGGGCAAG TCGGGGCAGG GCTACTACGT GGAGATGAC
C GTGGGCAGCC 09/471,669 201 GAGGGGCAAG TCGGGGCAGG GCTACTACGT GGAGATGAC
C GTGGGCAGCC

* *****

* *****
 * *****
 *
 *

EP 251
 G TAACTTTGCA
 09/471,669 251
 G TAACTTTGCA

CCCCGCAGAC GCTCAACATC CTGGTGGATA CAGGCAGCA
 CCCCCGAGAC GCTCAACATC CTGGTGGATA CAGGCAGCA

* *****
 * *****
 * *****
 * *****
 * *****

EP 301
 C AGAGGCAGCT
 09/471,669 301
 C AGAGGCAGCT

GTGGGTGCTG CCCCCACCC CTCCTGCAT CGCTACTAC
 GTGGGTGCTG CCCCCACCC CTCCTGCAT CGCTACTAC

* *****
 * *****
 * *****
 * *****
 * *****

EP 351 GTCCAGCACa TACCGGGACC TCCGGAAGGG TGTGTATG_a
G CCCTACACCC
09/471,669 351 GTCCAGCACa TACCGGGACC TCCGGAAGGG TGTGTATG_t
G CCCTACACCC

* *

* *

* *

* *

* *

EP 401 AGGGCAAGTG GGAAGGGGAG CTGGGCACCG ACCTGGTAA
G CATCCCCCAT
09/471,669 401 AGGGCAAGTG GGAAGGGGAG CTGGGCACCG ACCTGGTAA
G CATCCCCCAT

* *

* *

* *

* *

* *

EP 451 GGCCCCAACG TCACTGTGCG TGCCAACATT GCTGCCATC
A CTGAATCAGA
09/471,669 451 GGCCCCAACG TCACTGTGCG TGCCAACATT GCTGCCATC
A CTGAATCAGA

* *

* *

* *****
* *****
* *****

EP 501 CAAGTTCTTC ATCAACGGCT CCAACTGGGA AGGCATCCT
G GGGCTGGCCT
09/471,669 501 CAAGTTCTTC ATCAACGGCT CCAACTGGGA AGGCATCCT
G GGGCTGGCCT

* *****
* *****
* *****
* *****
* *****

EP 551 ATGCTGAGAT TGCCAGGCCT GACGACTCCC TGGAGCCTT
T CTTTGA CTCT
09/471,669 551 ATGCTGAGAT TGCCAGGCCT GACGACTCCC TGGAGCCTT
T CTTTGA CTCT

* *****
* *****
* *****
* *****
* *****

EP 601 CTGGTAAAGC AGACCCACGT TCCCAACCTC TTCTCCCTG
C AGCTTTGTGG

09/471,669 601
C AGCTTTGTGG

CTGGTAAAGC AGACCCACGT TCCCAACCTC TTCTCCCTG

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

EP 651
T GTCGGAGGGA
09/471,669 651
T GTCGGAGGGA

TGCTGGCTTC CCCCTCAACC AGTCTGAAGT GCTGGCCTC
TGCTGGCTTC CCCCTCAACC AGTCTGAAGT GCTGGCCTC

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

EP 701
G CAGTCTCTGG
09/471,669 701
G CAGTCTCTGG

GCATGATCAT TGGAGGTATC GACCACTCGC TGTACACAG
GCATGATCAT TGGAGGTATC GACCACTCGC TGTACACAG

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

* ****

* ****

EP 751

A TTGTGCGGGT

09/471,669 751

A TTGTGCGGGT

TATACACCCA TCCGGCGGGA GTGGTATTAT GAGGTGATC

TATACACCCA TCCGGCGGGA GTGGTATTAT GAGGTGATC

* ****

* ****

* ****

* ****

* ****

EP 801

G TACAACATG

09/471,669 801

G TACAACATG

GGAGATCAAT GGACAGGATC TGAAAATGGA CTGCAAGGA

GGAGATCAAT GGACAGGATC TGAAAATGGA CTGCAAGGA

* ****

* ****

* ****

* ****

* ****

EP 851

T GCCCAAGAAA

09/471,669 851

T GCCCAAGAAA

ACAAGAGCAT TGTGGACAGT GGCACCACCA ACCTTCGTT

ACAAGAGCAT TGTGGACAGT GGCACCACCA ACCTTCGTT


```
* ****
* ****
* ****
* ****
* ****

*****
*****
*****
*****
*****
*****
*****
```

EP 901
T CCACGGAGAA
09/471,669 901
T CCACGGAGAA

GTGTTTGAAG CTGCAGTCAA ATCCATCAAG GCAGCCTCC
GTGTTTGAAG CTGCAGTCAA ATCCATCAAG GCAGCCTCC

```
* ****
* ****
* ****
* ****
* ****

*****
*****
*****
*****
*****
*****
*****
```

EP 951
C TGGCAAGCAG
09/471,669 951
C TGGCAAGCAG

GTTCCCTGAT GGTTTCTGGC TAGGAGAGCA GCTGGTGTG
GTTCCCTGAT GGTTTCTGGC TAGGAGAGCA GCTGGTGTG

```
* ****
* ****
* ****
* ****

*****
*****
*****
*****
*****
```

* ****

EP 1001
A CCTAATGGGT
09/471,669 1001
A CCTAATGGGT

GCACCACCCC TTGGAACATT TTCCCAGTCA TCTCACTCT
GCACCACCCC TTGGAACATT TTCCCAGTCA TCTCACTCT

* ****

* ****

* ****

* ****

* ****

EP 1051
C AGCAATACCT
09/471,669 1051
C AGCAATACCT

GAGGTTACCA ACCAGTCCTT CCGCATCACC ATCCTTCCG
GAGGTTACCA ACCAGTCCTT CCGCATCACC ATCCTTCCG

* ****

* ****

* ****

* ****

* ****

EP 1101
T TACAAGTTTG
09/471,669 1101
T TACAAGTTTG

GCGGCCAGTG GAAGATGTGG CCACGTCCCA AGACGACTG
GCGGCCAGTG GAAGATGTGG CCACGTCCCA AGACGACTG

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

EP 1151
T TATCATGGAG
09/471,669 1151
T TATCATGGAG

CCATCTCACA GTCATCCACG GGCAGTGTTA TGGGAGCTG
CCATCTCACA GTCATCCACG GGCAGTGTTA TGGGAGCTG

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

EP 1201
G GCTTTGCTGT
09/471,669 1201
G GCTTTGCTGT

GGCTTCTACG TTGTCTTTGA TCGGGCCCGA AAACGAATT
GGCTTCTACG TTGTCTTTGA TCGGGCCCGA AAACGAATT

```
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
```

EP 1251
G GTGGAAGGCC
09/471,669 1251
G GTGGAAGGCC

CAGCGCTTGC CATGTGCACG ATGAGTTCAG GACGGCAGC
CAGCGCTTGC CATGTGCACG ATGAGTTCAG GACGGCAGC

* *****
* *****
* *****
* *****
* *****
* *****

EP 1301
T TCCACAGACA
09/471,669 1301
T TCCACAGACA

CTTTTGTAC CTTGGACATG GAAGACTGTG GCTACAACA
CTTTTGTAC CTTGGACATG GAAGACTGTG GCTACAACA

* *****
* *****
* *****
* *****
* *****
* *****

EP 1351
G CCATCTGCGC
09/471,669 1351
G CCATCTGCGC

GATGAGTCAA CCCTCATGAC CATAGCCTAT GTCATGGCT
GATGAGTCAA CCCTCATGAC CATAGCCTAT GTCATGGCT

* *****

* ****
* ****
* ****
* ****

EP 1401
G CGCTGCCTCC
09/471,669 1401
G CGCTGCCTCC

CCTCTTCATG CTGCCACTCT GCCTCATGGT GTGTCAGTG
CCTCTTCATG CTGCCACTCT GCCTCATGGT GTGTCAGTG

* ****
* ****
* ****
* ****
* ****

EP 1451
T CTCCCTGCTG
09/471,669 1451
T CTCCCTGCTG

GCTGCCTGCG CCAGCAGCAT GATGACTTTG CTGATGACA
GCTGCCTGCG CCAGCAGCAT GATGACTTTG CTGATGACA

* ****
* ****
* ****
* ****

EP 1501 AAGtgaggag gcccatggga gaaagataga gattcccct
g ggaccacacc
09/471,669 1501 AAG-----

EP 1551 tccgtgggttc actttggtca caagtaggag acacagatg
g cacctgtggc
09/471,669 1504 -----

EP 1601 cagagcacct caggaccctc cccacccacc aaatgcctc
t gccttgatgg
09/471,669 1504 -----

EP 1651 agaaggaaaa ggctggcaag gtgggttcca gggactgta
c ctgtaggaaa
09/471,669 1504 -----

EP 1701 cagaaaagag aagaaagaag cactctgctg gcgggaata
c tcttggtcac
09/471,669 1504 -----

EP 1751 ctcaaattta agtcgggaaa ttctgctgct tgaaacttc
a gccctgaacc

09/471,669 1504

```
EP          1801   ttgtccacc attcctttaa attctccaac ccaaagtat
t cttcttttct
09/471,669 1504   -----
```

```
EP          1851   tagtttcaga agtactggca tcacacgcag gttaccttg
g cgtgtgtccc
09/471,669 1504   -----
```


EP 1901 tgtggtaccc gggcagagaa gagaccaagc ttgtttccc
t gctggccaaa
09/471,669 1504

EP 1951 gtcagtagga gaggatgcac agtttgctat ttgctttag
a gacagggact
09/471,669 1504

EP 2001 gtataaacia gcctaacatt ggtgcaaaga ttgcctctt
g aattaaiaaaa
09/471,669 1504

EP 2051 aaaaactaga ttgactattt atacaaatgg gggcggctg
g aaagaggaga
09/471,669 1504 -----

EP 2101 aggagagggg gtacaaagac aggggaatagt gggatcaaa
g ctaggaaagg
09/471,669 1504 -----

EP 2151 cagaaacaca accactcacc agtcctagtt ttagacctc
a tctccaagat
09/471,669 1504

EP 2201 agcatcccat ctcagaagat ggggtggtggt ttcaatggt
t tctttttctgt
09/471,669 1504

EP 2251 ggttgcagcc tgaccaaaag tgagatggga agggcttat
c tagccaaaga
09/471,669 1504

```
EP          2301  gctctttttt agctctctta aatgaagtgc ccactaagg
a agttccactt
09/471,669 1504  -----
```

```
EP          2351  gaacacatgg aatttctgcc atattaattt ccattgtct
c tatctggaac
09/471,669 1504  -----
```

EP 2401 caccctttaa tctctacata tgattaggto cagcacttg
a aaatattcct
09/471,669 1504 -----

EP 2451 aaccnnaatt tgncttgggg gctttgcngn ccagggtgct
a aaagggnttg
09/471,669 1504 -----

EP 2501 ggtaggngnc cnccttntatn tnatncctna aaaggttan
n g
09/471,669 1504 -----

Alignment (FASTA format):

=====

>EP

ATGGCCCAAGCCCTGCCCTGGCTCCTGCTGTGGATGGGCGCGGGAGTGCT
GCCTGCCCACGGCACCCAGCACGGCATCCGGCTGCCCCTGCGCAGCGGCC
TGGGGGGCGCCCCCTGGGGCTGCGGCTGCCCCGGGAGACCGACGAAGAG
CCCGAGGAGCCCGGCCGGAGGGGCGAGCTTTGTGGAGATGGTGGACAACCT
GAGGGGCAAGTCGGGGCAGGGCTACTACGTGGAGATGACCGTGGGCAGCC
CCCCGCAGACGCTCAACATCCTGGTGGATACAGGCAGCAGTAACTTTGCA
GTGGGTGCTGCCCCCACCCTTCCCTGCATCGCTACTACCAGAGGCAGCT
GTCCAGCACATAACGGGACCTCCGGAAGGGTGTGTATGAGCCCTACACCC
AGGGCAAGTGGGAAGGGGAGCTGGGCACCGACCTGGTAAGCATCCCCAT
GGCCCCAACGTCACTGTGCGTGCCAACATTGCTGCCATCACTGAATCAGA
CAAGTTCTTCATCAACGGCTCCAACCTGGGAAGGCATCCTGGGGCTGGCCT
ATGCTGAGATTGCCAGGCCTGACGACTCCCTGGAGCCTTTCTTTGACTCT
CTGGTAAAGCAGACCCACGTTCCCAACCTCTTCTCCCTGCAGCTTTGTGG
TGCTGGCTTCCCCCTCAACCAGTCTGAAGTGCTGGCCTCTGTGCGGAGGGA
GCATGATCATTGGAGGTATCGACCACTCGCTGTACACAGGCAGTCTCTGG
TATACACCCATCCGGCGGGAGTGGTATTATGAGGTGATCATTGTGCGGGT
GGAGATCAATGGACAGGATCTGAAAATGGACTGCAAGGAGTACAACTATG
ACAAGAGCATTGTGGACAGTGGCACCACCAACCTTCGTTTGGCCCAAGAAA
GTGTTTGAAGCTGCAGTCAAATCCATCAAGGCAGCCTCCTCCACGGAGAA
GTTCCCTGATGGTTTCTGGCTAGGAGAGCAGCTGGTGTGCTGGCAAGCAG
GCACCACCCCTTGGAACATTTTCCAGTCATCTCACTCTACCTAATGGGT
GAGGTTACCAACCAGTCCTTCCGCATCACCATCCTTCCGCAGCAATACCT
GCGGCCAGTGGGAAGATGTGGCCACGTCCCAAGACGACTGTTACAAGTTTG
CCATCTCACAGTCATCCACGGGCACTGTTATGGGAGCTGTTATCATGGAG
GGCTTCTACGTTGTCTTTGATCGGGCCCCGAAAACGAATTGGCTTTGCTGT
CAGCGCTTGCCATGTGCACGATGAGTTCAGGACGGCAGCGGTGGAAGGCC
CTTTTGTACCTTGGACATGGAAGACTGTGGCTACAACATTCCACAGACA
GATGAGTCAACCCTCATGACCATAGCCTATGTCATGGCTGCCATCTGCGC
CCTCTTCATGCTGCCACTCTGCCTCATGGTGTGTGAGTGGCGCTGCCTCC
GCTGCCTGCGCCAGCAGCATGATGACTTTGCTGATGACATCTCCCTGCTG

AAGtgaggaggcccatgggagaaagatagagattcccctgggaccacacc
tccgtgggttcactttgggtcacaagtaggagacacagatggcacotgtggc
cagagcacctcaggaccctccccaccaccaaatagcctctgccttgatgg
agaaggaaaaggctggcaagggtgggttccagggactgtacctgtaggaaa
cagaaaagagaagaaagaagcactctgctggcggaataactcttggtcac
ctcaaatttaagtgcgggaaattctgctgcttgaaacttcagccctgaacc
tttgtccaccattcctttaaattctccaacccaaagtattcttcttttct
tagtttcagaagtactggcatcacacgcagggttaccttggcgtgtgtccc
tgtgggtaccggggcagagaagagaccaagcttggttccctgctggccaaa
gtcagtaggagaggatgcacagtttgctatttgcttttagagacagggact
gtataaacaagcctaacattgggtgcaaagattgcctcttgaaattaaaaaa
aaaaactagattgactatttatacaaatggggggcggtggaaagaggaga
aggagaggggagtacaaagacagggaatagtgggatcaaagctaggaaagg
cagaaacacaaccactcaccagtcctagttttagacctcatctccaagat
agcatcccctctcagaagatgggtgtgttttcaatgttttcttttctgt
ggttgacgctgaccaaagtgagatgggaagggttatctagccaaaga
gctcttttttagctctcttaaataagtgccactaaggaagttccactt
gaacacatggaatttctgccaataattaatttccattgtctctatctggaac
caccctttaatctctacatatgattaggtccagcacttgaaaatatctct
aaccnnaatttgncttgggggctttgcngnccagggtgctaaaagggnttg
ggtaggngnecncttntatntnatnccnnaaaagggttanng

>09/471,669

ATGGCCCAAGCCCTGCCCTGGCTCCTGCTGTGGATGGGCGCGGGAGTGCT
GCCTGCCACGGCACCCAGCACGGCATCCGGCTGCCCTGCGCAGCGGCC
TGGGGGGCGCCCCCTGGGGCTGCGGCTGCCCGGGAGACCGACGAAGAG
CCCGAGGAGCCCGGCCGGAGGGGCAGCTTTGTGGAGATGGTGGACAACCT
GAGGGGCAAGTCGGGGCAGGGCTACTACGTGGAGATGACCGTGGGCAGCC
CCCCGCAGACGCTCAACATCCTGGTGGATACAGGCAGCAGTAACTTTGCA
GTGGGTGCTGCCCCCCCACCCCTTCCTGCATCGCTACTACCAGAGGCAGCT
GTCCAGCACATACCGGGACCTCCGGAAGGGTGTGTATGTGCCCTACACCC
AGGGCAAGTGGGAAGGGGAGCTGGGCACCGACCTGGTAAGCATCCCCCAT
GGCCCCAACGTCACTGTGCGTGCCAACATTGCTGCCATCACTGAATCAGA
CAAGTTCTTCATCAACGGCTCCAAGTGGGAAGGCATCCTGGGGCTGGCCT
ATGCTGAGATTGCCAGGCCTGACGACTCCCTGGAGCCTTTCTTTGACTCT
CTGGTAAAGCAGACCCACGTTCCCAACCTCTTCTCCCTGCAGCTTTGTGG
TGCTGGCTTCCCCCTCAACCAAGTCTGAAGTGCTGGCCTCTGTGCGAGGGA
GCATGATCATTGGAGGTATCGACCACTCGCTGTACACAGGCAGTCTCTGG
TATACACCCATCCGGCGGGAGTGGTATTATGAGGTGATCATTGTGCGGGT
GGAGATCAATGGACAGGATCTGAAAATGGACTGCAAGGAGTACAACTATG
ACAAGAGCATTGTGGACAGTGGCACCAACCAACCTTCGTTTGCCCAAGAAA
GTGTTTGAAGCTGCAGTCAAATCCATCAAGGCAGCCTCCTCCACGGAGAA
GTTCCCTGATGGTTTCTGGCTAGGAGAGCAGCTGGTGTGCTGGCAAGCAG
GCACCACCCCTTGGAACATTTTCCCAGTCATCTCACTCTACCTAATGGGT
GAGGTTACCAACCAGTCCTTCCGCATCACCATCCTTCCGCAGCAATACCT
GCGGCCAGTGGAAGATGTGGCCACGTCCCAAGACGACTGTTACAAGTTTG

CCATCTCACAGTCATCCACGGGCACTGTTATGGGAGCTGTTATCATGGAG
GGCTTCTACGTTGTCTTTGATCGGGCCCGAAAACGAATTGGCTTTGCTGT
CAGCGCTTGCCATGTGCACGATGAGTTCAGGACGGCAGCGGTGGAAGGCC
CTTTTGTACCTTGGACATGGAAGACTGTGGCTACAACATTCCACAGACA
GATGAGTCAACCCTCATGACCATAGCCTATGTCATGGCTGCCATCTGCGC
CCTCTTCATGCTGCCACTCTGCCTCATGGTGTGTCAGTGGCGCTGCCTCC
GCTGCCTGCGCCAGCAGCATGATGACTTTGCTGATGACATCTCCCTGCTG
AAG-----